[redacted]

September 18, 2018

Taxonomic Identification Report

Submission: J-18-0044 (same as R-15-0003)
Analyst: Carolina Peñalva-Arana, PhD
Submitter: National Institute of Standards and Technology
Material Measurement Laboratory
100 Bureau Drive
Gaithersburg, MD 20899-8312

Production Strain
There is one TERA strain *Saccharomyces cerevisiae* NE095 (NIST ERCC 00095)

Recipient strain
The recipient strain is *Saccharomyces cerevisiae* BY4739 (MATalpha leu2Δ0 lys2Δ0 ura3Δ0) derived from *S. cerevisiae* S288C, and was assessed by EPA under R-15-0003.

Parent Strain

***S. cerevisiae* S288C Lineage** (GenBank Taxonomy ID 559292)
Domain: Eukaryota
Kingdom: Fungi
Subkingdom: Dikarya
Phylum: Ascomycota
Subphylum: Saccharomycotina
Class: Saccharomycetes
Order: Saccharomycetales
Family: Saccharomycetaceae
Genus: *Saccharomyces*
Species: *cerevisiae*

Methods for confirming taxonomic identity

S. cerevisiae BY4739 strain was derived from S288C by Prof. Jef Boeke's laboratory at John Hopkins University as described by Brachmann et al. (1998). Jef Boeke then provided his strain

to Open Biosystems (now Dharmacon GE Healthcare, product #YSC1062). NIST procured BY4739 from Open Biosystems. Open Biosystems (now Dharmacon GE Healthcare) lists the organism as product # YSC1062. It is deposited as well at the American Type Culture Collection (ATCC 200901). Strain YGSC S288C was original held in the Yeast Genetic Stock Center as shown by the Histri in Straininfo (Straininfo, 2015). YGSC has since been absorbed into the ATCC Fungi and Yeast collection (ATCC). *S. cerevisiae* strain S288C is a commonly used laboratory strain and was used in the yeast systematic sequencing project (SGD, 2015).

In addition, the taxonomy of the BY4739 strain was substantiated via comparison of the subject microorganism NE095, S288C, and the BY4739 strains using whole genome sequencing (WGS) data. The Ion Torrent PGM sequencing platform (Life Technologies Corporation) was used to generate the whole genome sequencing of the subject NE095 strain and it was then compared with the available S288C strain WGS. The submitters supplied BLAST statistics showing 99% identity of the subject microorganism, the recipient strain and S288C.

Visual inspection of WGS alignments confirmed gene deletions in the BY4739 strain, relative to the S288C, and confirmed the absence of LEU2 and LYS2 genes in the recipient strain NE095. These results substantiate the taxonomy of the recipient microorganism as BY4739 derived from S288C. Additional whole genome sequencing with the Illumina MiSeq platform was performed at a later date by the submitter with another batch of the material and the MiSeq data further substantiated the taxonomic findings.

The subject microorganism, *S. cerevisiae* NE095, is taxonomically identical to the recipient strain, with the main differences being the insertion of the URA3 gene and the ERCC-00095 sequence, which is will be used as a selection/detection marker, the sole purpose for this strain.

Donor organism

***M. jannaschii* DSM 2661 Lineage** (GenBank Taxonomy ID 243232)
Domain: Archaea
Kingdom: Euryarchaeota
Phylum: Euryarchaeota
Class: Methanococci
Order: Methanococcales
Family: Methanocaldococcaceae

The introduced marker sequence was obtained from *Methanocaldococcus jannaschii* DSM 2661. The donor microorganism is an autotrophic hyperthermophilic methanogenic archaea found near hypothermal vents in the ocean. *M. jannaschii* DSM 2661 has been fully sequenced (GenBank GCA_000091665.1). The insert intergeneric sequence was reconstructed from the

pYES2 and ERCC-00095 plasmids *in-silico.* The intergeneric DNA was determined to have been inserted in chromosome IV of the subject microorganism and includes a 1106 bp sequence containing the *S. cerevisiae* URA3 gene (NCBI Gene ID 856692) and a 438 bp sequence from ERCC-00095 (GenBank Accession KC702204). The URA3 gene confers the ability for the subject yeast to grow in yeast synthetic defined (SD) broth without uracil. A previous synomym of M*ethanocaldococcus* is *Methanococcus*.

Intergeneric sequences

URA3 gene - The subject microorganism has a functional URA3 gene, which enables selection of successfully transformed cells via growth in uracil-deficient medium. Thus, the subject microorganism can grow in both typical yeast broth (such as yeast extract peptone dextrose, YPD) or in yeast synthetic defined (SD) broth without uracil (i.e., Clontech 630314 SD/-Ura Broth). Otherwise, the subject microorganism is not anticipated to have any other changes in phenotype or traits as compared to the recipient microorganism.

ERCC-00095 gene
ERCC-00095 (GenBank Accession KC702204) was selected as a specific selection marker for the subject microorganism. ERCC-00095 is part of the NIST Standard Reference Material (SRM) 2374 Sequence Library for External RNA Controls. It is derived from *M. jannaschii* spike-in control MJ-500-42 genomic sequence (GenBank: DQ516759.1) from *M. jannaschii* DSM 2661.
The corresponding protein for this sequence is phosphate specific transport complex component (pstB, GenBank: AAB99016.1).

ERCC-00095 only includes the final three out of eight open reading frames in the protein and 62 out of the 252 amino acids; thus, the sequence is not expected to result in a functional protein. In the unlikely event that it were expressed, it should not confer new metabolic capabilities due to three primary reasons: (1) only the latter two-thirds of the gene was incorporated into the genome, such that the full protein cannot be expressed; (2) the gene is part of a phosphate transport mechanism and without the other genes involved in the metabolic pathway it is unlikely to function; and (3) the recipient microorganism already possesses a phosphate transport system, and therefore if the gene were somehow expressed and functional it would not provide the subject microorganism with a new metabolic capability.

References

American Type Culture Collection. Fungi and Yeast. www.atcc.org/Products/Cells and Microorganisms/Fungi and Yeast.aspx

C. Kurtzman and C. Robnett (1998). Identification and phylogeny of ascomycetous yeasts from

analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie van Leeuwenhoek 73: 331-371.

Straininfo. Strain Passport for S288C *Saccharomyces cerevisiae.* www.straininfo.net/strains/317495

Saccharomyces Genome Database (SGD). www.yeastgenome.org

Brachmann C, A. Davies, G. Cost, E. Caputo, J. Li, P. Hieter, and J. Boeke. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast., **14**:115-132.

Segal, Mark. 2015. Taxonimic ID report for R-15-0003. USEPA.